Article

Phenotypic Correlation Analysis in F₂ Segregating Populations of *Gossypium hirsutum* and *Gossypium arboreum* for Boll-Related Traits

Zubair Iqbal ¹, Daowu Hu ¹, Wajad Nazeer ², Hao Ge ¹, Talha Nazir ³, Sajid Fiaz ⁴, Alia Gul ⁵, Muhammad Shahid Iqbal ¹, Ahmed M. El-Sabrou ⁶, Zahra Maryum ⁷, Zhaoe Pan ¹ and Xiongming Du ^{1,*}

- ¹ State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China; zubairzube@gmail.com (Z.I.); 15637266573@163.com (D.H.); gehao202024@163.com (H.G.); shahidkooia@gmail.com (M.S.I.); panzhaoe@163.com (Z.P.)
 - ² Department of Plant Breeding and Genetics, Faculty of Agricultural Sciences, Ghazi University, Dera Ghazi Khan 32200, Punjab, Pakistan; wnazeer@gudgk.edu.pk
 - ³ Department of Plant Protection, Faculty of Agricultural Sciences, Ghazi University, Dera Ghazi Khan 32200, Punjab, Pakistan; talha23december@gmail.com
 - ⁴ Department of Plant Breeding and Genetics, The University of Haripur, Haripur 22620, Khyber Pakhtunkhwa, Pakistan; sfiaz@uoh.edu.pk
 - ⁵ Department of Botany, Hazara University, Mansehra 21120, Khyber Pakhtunkhwa, Pakistan; aliagulbotanist@gmail.com
 - ⁶ Department of Applied Entomology and Zoology, Faculty of Agriculture (EL-Shatby), Alexandria University, Alexandria 21545, Egypt; elsabroutahmed@alexu.edu.eg
 - ⁷ Cotton Laboratory, Plant Breeding and Genetics Division, Nuclear Institute for Agriculture and Biology (NIAB), Faisalabad 38000, Punjab, Pakistan; zahramaryum2@gmail.com
- * Correspondence: duxiongming@caas.cn

Citation: Iqbal, Z.; Hu, D.; Nazeer, W.; Ge, H.; Nazir, T.; Fiaz, S.; Gul, A.; Iqbal, M.S.; El-Sabrou, A.M.; Maryum, Z.; et al. Phenotypic Correlation Analysis in F₂ Segregating Populations of *Gossypium hirsutum* and *Gossypium arboreum* for Boll-Related Traits. *Agronomy* **2022**, *12*, 330. <https://doi.org/10.3390/agronomy12020330>

Academic Editor: José Iñaki Hormaza

Received: 18 December 2021

Accepted: 24 January 2022

Published: 27 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Cotton is an important agro-industrial crop across the globe. Improving the fiber quality and yield potential of cotton are major commercial targets for cotton breeders. The cotton lint yield is computed by multiplying three fundamental yield constituents: average boll weight, boll number per unit ground area, and lint percentage. The cotton species *Gossypium arboreum* exhibits a wide range of desirable traits, which are absent in the congener *Gossypium hirsutum*. Four parental lines of *G. hirsutum* and *G. arboreum*, with significant differences in boll-related traits, were used to develop the following four F₂ populations: Mei Zhongmian × Chimu Heizi (MC), Mei Zhongmian × L-02292-3 (ML), Dixie king × Suyuan 04-44 (DS), and Dixie king × Pamuk (DP), in order to study complex traits, such as boll weight (BW) (g), lint percentage (LP) (%), boll upper width (BUW), boll medium width (BMW), boll lower width (BLU), and boll length (BL) (mm). In segregation populations, extensive phenotypic differences and transgressive segregation were observed. The results show that most of the correlation clusters were negatively associated with boll weight and lint percentage. The positive correlation clusters were observed among boll upper width (BUW), boll medium width (BMW), boll lower width (BLW), and boll length (BL). Seven of the twenty-four extracted principal components had eigenvalues > 1. This accounted for 62.2% of the difference between the four F₂ populations. Principal component 1 accounted for 15.1% of the overall variability. The variation in principal component 1 was mainly attributed to boll lower width (BLW), boll medium width (BMW), boll upper width (BUW), boll length (BL), and boll weight (BW) of the ML population. The heritability estimates varied between high, medium, and low for various traits among the studied F₂ populations. Interestingly, all traits demonstrated low genetic advance, which indicates that non-additive genes controlled these characters and that direct selection for these traits is not beneficial. The outcome of the present investigation will help to develop cotton cultivars with improved boll weight and lint percentage.

Keywords: *Gossypium hirsutum*; *Gossypium arboreum*; intraspecific hybridization; genetic resource; yield improvement

1. Introduction

Cotton is a globally valuable, synthetic cash crop known for its natural fibers [1]. *Gossypium* is a genus with seven tetraploid ($2n = 4x = 52$) and forty-five diploid ($2n = 2x = 26$) species, such as *G. herbaceum* L. (A1), *G. arboreum* L. (A2), *G. hirsutum* L. (AD1), and *G. barbadense* L. (AD2) [2–4]. Upland cotton (*Gossypium hirsutum* L.) is a commonly cultivated cotton species, accounting for 95% of worldwide cotton production owing to its high yielding capacity [5,6]. The economic and environmental adaptability of *G. hirsutum* has been influenced by artificial breeding and long-term natural selection, and its production continues to increase globally across a wide range of latitudes [7]. Additionally, *Gossypium arboreum* L. was introduced from India and later domesticated and cultivated in China for nearly 2000 years. Currently, China is the world's largest cotton producer and raw cotton consumer, accounting for one-quarter of global cotton production and one-third of global cotton consumption [8]. The species *G. arboreum* exhibits a wide range of desirable traits for cotton cultivation, which are absent in upland cotton. The global significance of the cotton crop has prompted plant breeders to enhance the genetic infrastructure of the cotton plant [5]. Moreover, *G. arboreum* holds the ability to display drought resistance, tolerance to diseases, and is effectively adapted to low input and dry land conditions [9]. Therefore, *G. arboreum* can serve as a precious resource for future hybridization programs. The Cotton Germplasm Center at the Chinese Academy of Agricultural Science has collected and conserved over 300 *G. arboreum* accessions [10].

Through interspecific hybridization, the cotton crop has been improved in terms of both biotic and abiotic stress tolerance and fiber quality traits. Cotton breeders have performed significant crosses among wild accessions and upland cotton to produce offspring with high fiber content [11–13]. Furthermore, to meet the new challenges posed by the textile industry, genetic changes effecting cotton fiber yield and quality traits have been achieved through long-term selective and artificial breeding [14]. Improving cotton fiber production and potential productivity is a key commercial target for cotton breeders. Cotton lint yield is computed via the multiplication of three fundamental yield constituents: average BW, bolls number per unit ground area, and LP [15]. For cotton cultivars, LP is a significant yield-related attribute and a major economic indicator [16,17]. Furthermore, the main contributor for lint yield is boll number per unit land area, followed by seed number per boll and lint mass per seed [18]. Improved lint yield has been achieved with low seed counts, ensuring that a high percentage of lint can be produced when limited by seed number [16]. Cotton bolls can also be used as fiber estimation markers, in order to infer the genetic and physiological processes underlying yield and plant growth, since they are linked to productivity. Thus, the profiling of high-throughput sequencing data for cotton bolls plays an important role in advancing modern cultivars and the efficiency of management strategies. The conventional quantification approaches of cotton bolls are laborious and time-consuming. Therefore, new approaches can enhance the throughput phenotyping potential of cotton bolls [19].

Correlation analysis has shown positive associations between seed cotton yield, boll weight, BL, BCW, BD, and bolls per plant [20]. Moreover, boll number and size have a significant impact on the final cotton yield, whereas the shape of the boll determines its fiber quality. Cotton breeders are interested in using diploid genetic material to enhance the production of cultivated tetraploid cotton resistant to both biotic and abiotic stresses. On a global level, especially in China, breeders are working to develop cultivable diploid cotton with premium characteristics, i.e., BW, LP, and other important fiber quality features. Thus far, numerous cultivated diploid cotton accessions have improved and enhanced the invaluable genetic resources used in cotton biotechnological research. In the present study, we investigated two within-boll yield components, BW and LP, and four boll morphological traits, boll length, and upper, medium, and lower boll width, to examine the diversity among cotton cultivars produced through intraspecific hybridization.

2. Materials and Methods

2.1. Parental Plant Material and Hybridization

Two parental lines from *G. hirsutum* and *G. arboreum* were used to develop F₂ populations displaying major changes in boll-related traits. Selected accessions were preserved at China National GeneBank, Institute of Cotton Research, Chinese Academy of Agriculture Sciences, Anyang, Henan, China. Four crosses were developed, i.e., two crosses for *G. arboreum* (ID-MC and ML); two crosses for *G. hirsutum* (DS and DP). For all four crosses, lines with big bolls were used as female parents, and lines with small bolls were used as male parents. The list of crosses is shown in Table 1.

Table 1. List of parental lines for hybridization on the basis of boll weight (BW_g).

Genotypes	Parents	Species	BW (g)
Mei Zhongmian	Female	<i>G. arboreum</i>	3.76
Chimu Heizi	Female	<i>G. arboreum</i>	1.28
L-02292-3	Male	<i>G. arboreum</i>	1.33
Dixie king	Female	<i>G. hirsutum</i>	6.71
Suyuan 04-44	Male	<i>G. hirsutum</i>	3.71
Pamuk	Male	<i>G. hirsutum</i>	3.75

2.2. Planting and Phenotyping

The parental genotypes of *G. hirsutum* and *G. arboreum*, and their F₂ populations, were sown at the Institute of Cotton Research in Anyang, Henan, China, between mid-April and late October, the normal cotton growing season (near yellow River), during years 2019 and 2020. The longitude and latitude coordinates of the site are 114.07° E and 35.85° N, respectively. In the experimental area, all accessions were cultivated in three replicates using a randomized complete block pattern. Each entry plot was 7 × 3 m in size, with row-to-row and plant-to-plant distances of 76 and 30 cm, respectively. Normal agronomic practices were undertaken and same scoring parameters were utilized in both years for the collection of phenotypic data.

2.3. Sample Preparation for Hybridizations

We investigated six traits, four agronomic boll morphological traits and two economic yield-related traits, of four cotton species, including two parents of *G. hirsutum* and two parents of *G. arboreum*, the sexes of which were based on boll weight; other associated traits were used for the evaluation of F₂ generations developed through intraspecific hybridization. The female parents were superior in their characteristics, i.e., BW, BL, and LP. The male parents possessed extremely low BW, LP, and morphological traits (Tables 1 and 2). The phenotypic data for traits under investigation were recorded upon reaching harvest level.

Table 2. Intraspecific hybridization block.

Genotypes	ID	Crosses	Species	BW (g)
Mei Zhongmian × Chimu Heizi	MC	Big boll × Small boll	<i>G. arboreum</i>	2.1
Mei Zhongmian × L-02292-3	ML	Big boll × Small boll	<i>G. arboreum</i>	3.0
Dixie king × Suyuan 04-44	DS	Big boll × Small boll	<i>G. hirsutum</i>	6.2
Dixie king × Pamuk	DP	Big boll × Small boll	<i>G. hirsutum</i>	5.0

2.4. Measurement of Traits

Six plants were selected for each replication to evaluate the following boll-related traits: BW (g); BUW, BUM, and BLW (mm); top-to-bottom length of cotton boll (BL, mm); LP (%). The weighed bolls were collected from central fruiting branches.

2.5. Statistical Analysis

To compute the correlation of traits under observation, XLSTAT and “R” software (package “corrplot”) were used. Minitab 18 was used to conduct the principal component analysis (PCA). The Dewey and Lu formula was used to calculate correlation coefficients [21]. Adopting the method of Fruchterman and Reingold [22], we generated a force-directed design by following the algorithm for the distance among traits, from which we derived proportional absolute correlation values. Finally, positive correlations were illustrated in pink, while negative correlations were illustrated in blue. Thicker lines show stronger correlations among traits, whereas thinner lines represent weaker correlations among the traits. The correlation network procedure was undertaken using Cytoscape bioinformatics software.

3. Results

3.1. Variation in Phenotypic Traits of Four F_2 Populations

We investigated four crosses, i.e., two crosses for *G. arboreum* (ID-MC and ML) and two for *G. hirsutum* (DS and DP). For all four crosses, lines with large boll sizes were used as female parents, and lines with small boll sizes were considered as male parents. In *G. hirsutum*, small boll parents were nearly half the weight of lines with larger bolls; in *G. arboreum*, the boll weights of small boll parents were nearly one third of the large boll lines. Thus, the F_2 populations exhibited more diversity in boll weight.

Numerous divergent phenotypes and transgressive segregants were noticed in the F_2 populations (Table 3). Transgressive segregation occurs when certain individuals of the next generation have phenotypic values greater than those of their superior parent [23]. The coefficients of variation (CV) indicated different levels of variability for the six traits, i.e., LP, BW, BUW, BMW, BLW, and BL (Table 3). The CV values for LP were low (*G. arboreum*: 7.84–11.18%, *G. hirsutum*: 9.02–9.29%), whereas BW had high (*G. arboreum*: 17.18–21.16%, *G. hirsutum*: 15.24–16.54%) CV values. Among the traits BUW, BMW, BLW, and BL, the CV values were lowest for BLW (6.78–7.79%) in *G. hirsutum* and highest for BUW (10.91–9.14%) in *G. arboreum*. On comparing the *G. arboreum* and *G. hirsutum* populations, it was observed in *G. arboreum* populations that BW had the highest CV values, 21.67 and 17.17, for both ML and MC crosses, respectively. Similarly, in *G. hirsutum*, BW had the highest CV values, 16.54 and 15.24, for both DS and DP, respectively. Moreover, it was also recorded that crosses derived from *G. arboreum* had higher CV values than those from *G. hirsutum*.

3.2. Correlation Analysis

Correlation analysis revealed a significant association among the six traits across the four F_2 populations (Figure 1). In the ML population, BW exhibited increasingly significant positive correlations with the LP, BUW, BMW, BLW, and BL, respectively. Moreover, correlation analysis showed a strong positive association between the LP and BUW of *G. hirsutum*, and a significant negative correlation of these traits with the BL of the MC population. The BW in the MC population exhibited a significant strong positive correlation with the BUW, BMW, BLW, and BL. The BW of DP population showed a significant positive correlation with BUW, BMW, BLW, and BL, and a significant negative correlation with LP. The BW in the DS population showed a significant positive correlation with the BUW, BMW, BLW, and BL, and a significant negative correlation with the trait LP of same population, and with BUW of the ML population.

The LP of the ML population showed increasingly significant positive correlations with the BUW, BMW, BLW, and BL, respectively. Moreover, it showed a significant positive correlation with the BUW of the DP population. The LP of MC population showed a significant positive correlation with the BUW of the DS population. The LP of the DP population revealed a significant negative correlation with the BL of the MC population. The LP of the DS population revealed a strong negative association with BLW and BL, and had a significant negative correlation with BUW of the MC population. The BUW of the ML population showed increasingly significant positive associations with BMW, BLW,

and BL, respectively, whereas, it exhibited increasingly significant positive correlations with BUW of the MC, DP, and DS populations, respectively. It also revealed a significant negative association with BL of the MC population. The BUW of the ML population correlated positively with the BMW, BLW, and BL, with respectively increasing correlation strength. It also showed a significant positive correlation with BUW of the MC, DS, and DP populations, and a significant negative correlation with the BL of the MC population. The BUW of the MC population showed a significant positive correlation with BMW, BLW, BL, and BUW of the DP population. The BUW of the DP population showed a significant positive correlation with BMW, BLW, and BL of the same population, whereas it showed a significant positive correlation with the BMW and BLW of the DS population. The BUW of DS population showed significant positive correlations of increasing strength with the BMW, BLW, and BL, respectively. However, it also showed a significant positive correlation with BMW, BLW, and BL of ML population.

Table 3. Six traits demonstrate phenotypic variation in four F₂ populations: ML, MC, DS, and DP.

Population	Traits	P ₁	P ₂	F ₂ Populations				
				Min.	Max.	Mean	SD	CV%
ML	BLW	29.72	22.76	20.10	36.52	25.30	2.19	8.67
	BMW	29.11	22.61	17.49	33.47	24.02	2.17	9.07
	BUW	27.35	17.03	13.97	29.55	19.28	1.81	9.41
	BW	3.93	1.63	0.99	5.98	2.29	0.48	21.16
	LP	32.68	18.13	16.67	35.32	26.15	2.92	11.18
	BL	37.66	29.26	24.74	52.77	31.73	3.041	9.58
MC	BLW	29.72	22.52	17.99	29.48	23.68	1.94	8.22
	BMW	29.11	22.4	17.48	28.82	22.38	1.91	8.56
	BUW	27.35	17.52	12.94	24.51	17.68	1.93	10.91
	BW	3.93	1.7	1.38	4.05	2.26	0.38	17.18
	LP	32.68	23.87	20.47	40.29	28.81	2.26	7.84
	BL	37.66	27.01	20.12	37.66	30.03	3.01	10.03
DS	BLW	34.48	32.31	24.08	44.42	35.26	2.39	6.78
	BMW	34.03	31.13	26.09	39.43	32.96	2.24	6.79
	BUW	26.74	23.69	17.45	30.10	23.20	2.15	9.27
	BW	5.79	3.93	1.25	7.45	5.02	0.83	16.54
	LP	35.77	35.37	13.57	46.4	33.38	3.01	9.02
	BL	45.52	50.51	38.74	60.27	50.03	3.78	7.55
DP	BLW	34.48	26.35	24.02	37.97	30.78	2.39	7.79
	BMW	34.03	26	22.81	35.65	28.67	2.33	8.14
	BUW	26.74	21.34	15.25	31.33	20.95	2.12	10.12
	BW	5.79	3.65	1.6	6.39	4.40	0.67	15.24
	LP	35.77	31.81	13.57	46.46	32.39	3.01	9.29
	BL	45.52	36.81	30.14	54.26	42.14	3.69	8.77

P₁, parent 1; P₂, parent 2; Min, minimum; Max, maximum; SD, standard deviation; CV, coefficient of variation; ML, Mei Zhongmian × L-02292-3; MC, Mei Zhongmian × Chimu Heizi; DS, Dixie king × Suyuan 04-44; DP, Dixie king × Pamuk; BLW, boll lower width (mm); BMW, boll medium width (mm); BUW, boll upper width (mm); BW, boll weight (g); LP, lint percentage (%); BL, boll length (mm).

Figure 1. Frequency distribution of phenotypic variance and correlation coefficients for six boll-related traits in four F_2 populations: MC and ML of *G. arborium*; DS and DP of *G. hirsutum*. LP, lint percentage; BW, boll weight; BUW, boll upper width; BMW, boll medium width; BLW, boll lower width; BL, boll length. Significance is at $p = 0.05$, $p = 0.01$, $p = 0.001$, respectively.

The BMW of the ML population correlated positively with BLW and BL. The BMW of the MC population showed increasingly positive correlations with BLW and BL, respectively. Furthermore, it also revealed a significant positive correlation with BL of the ML population. The BMW of the DP population correlated positively with BLW and BL. The BMW of the DS population showed a significant positive correlation with BLW and BL. In addition, it also showed a significant positive correlation with BL of the DS population. The BLW of the ML population displayed a significant positive correlation with BL. Moreover, it also showed a significant negative correlation with BL of the MC population. The BLW of the DP population was found to be significantly positively correlated with BL. The BLW of the DS population displayed a significant positive correlation with BL. Moreover, it showed a significant negative correlation with BL of both the ML and DP populations. The network was constructed using phenotypic pairwise correlations between BW, LP, BUW, BMW, BLW, and BL among the four populations MC, ML, DS, and DP, respectively (Figure 2). Most of the correlation clusters were negatively associated with the two economic traits, i.e., BW and LP, whereas correlation clusters were found to be positively correlated with the majority of the agronomic traits under investigation. Significant correlation clusters were observed among multiple characters, i.e., BLW, BMW, and BUW of the ML population, BMW and BLW of the DS population displayed significant negative correlations with DP and BW of the DP population, BLW of the MC population, BW, LP, and BUW of the DS population, and BL of the MC, BUW of the ML, and BL of the MC populations, respectively.

Figure 2. Phenotypic correlation networks among four F₂ populations: MC and ML of *G. arborium*; DS and DP of *G. hirsutum*. Correlations are denoted by blue and pink lines that represent negative and positive correlations, respectively. The width of the line is proportional to correlation strength. BW, boll weight; BL, boll length; LP, lint percentage; BUW, boll upper width; BMW, boll medium width; BLW, boll lower width.

3.3. Estimation of Heritability

Variations among both phenotypic and genotypic characters were exhibited by the F_2 populations under investigation. The phenotypic and genotypic coefficient of variation (PCV and GCV), broad sense heritability (h^2b) of four agronomic (boll morphological), and two economic (yield-related) traits are summarized in Table 4. The range of GCV was between 4.40 and 19.62 for the six traits measured in all four populations. The highest value of GCV (19.62) was observed for BW of the ML population, followed by BW of the MC population (16.38), in *G. arboreum*, whereas in *G. hirsutum*, populations DS and DP displayed GCV values of 15.40 and 14.14, respectively. The range of PCV was 8.40 to 23.95 for all traits measured in all four populations. In addition, at the phenotypic level, BW revealed more variation in PCV, which suggested that all investigated populations have high genetic potential, which is critical for heterosis breeding programs because the transmissibility index (heritability) is restricted to the next generation. The majority of traits exhibited high h^2b , ranging by 15% for BMW and BUW of both the DP and MC populations, respectively. Estimations of h^2b were in a relatively medium range (high > 50%, medium 20–50%, low < 20%) for traits under investigation. High heritability results were proven to be effective in selecting superior genotypes based on phenotypic results. The highest genetic improvement value was related to LP (4.03–4.86) for all populations, followed by BL (4.37–2.67), and BLW and BMW (2.74 and 2.72) in the ML population, respectively. It was demonstrated that traits BW, LP, and BL hold high heritability and genetic improvement.

Table 4. Estimation of genetic variability and heritability parameters in ML, MC, DS, and DP F_2 populations.

Populaiton	Traits	GCV	PCV	h^2	GA
MC	BUW	6.45	16.57	15.13	0.91
	BMW	5.87	12.29	22.80	1.29
	BLW	5.44	11.99	20.62	1.21
	BL	7.66	13.60	31.73	2.67
	BW	16.38	18.69	76.78	0.67
	LP	7.56	8.40	80.86	4.03
ML	BUW	7.70	12.13	40.27	1.94
	BMW	7.84	11.15	49.44	2.73
	BLW	7.49	10.65	49.46	2.75
	BL	8.66	11.21	59.67	4.37
	BW	19.63	23.96	67.12	0.76
	LP	10.56	12.35	73.09	4.86
DP	BUW	6.18	15.21	16.53	1.08
	BMW	4.80	12.36	15.05	1.10
	BLW	4.95	11.55	18.33	1.34
	BL	6.67	11.92	31.33	3.24
	BW	14.15	17.24	67.34	1.05
	LP	8.66	10.46	68.48	4.78
DS	BUW	6.70	12.96	26.71	1.66
	BMW	4.71	9.71	23.46	1.55
	BLW	4.40	9.97	19.49	1.41
	BL	5.66	10.36	29.84	3.19
	BW	15.41	18.61	68.58	1.32
	LP	8.36	10.22	66.86	4.70

ML, Mei Zhongmian $\times$ L-02292-3; MC, Mei Zhongmian $\times$ Chimu Heizi; DS, Dixie king $\times$ Suyuan 04-44; DP, Dixie king $\times$ Pamuk; BLW, boll lower width; BMW, boll medium width; BUW, boll upper width; BW, boll weight; LP, lint percentage; BL, boll length; PCV, phenotypic coefficient variance; GCV, genotypic coefficient variance; h^2 , heritability; GA, genetic advance.

3.4. Principal Component Analysis (PCA)

Principal component analysis (PCA) was conducted on all variables at the same time to identify patterns of variance. The eigenvalues, variability (%), and cumulative percentages

(%) are presented in Table 5. Seven of the twenty-four extracted PCs had eigenvalues > 1, and accounted for 62.2% of the difference between the four populations. PC1 accounted for 15.1% of the overall variability and was mainly attributed to BLW, BMW, BUW, BL, and BW of the ML population of *G. arboreum*. PC2 accounted for 12.2% of the overall variability and contributed mostly negative values to BLW, BMW, BL, BUW, and BW of the DP population of *G. hirsutum*. PC3 accounted for 11.2% of the total variability and contributed negative loadings to BLW, BMW, and BUW of the MC population of *G. arboreum*. PC4 accounted for 9.6% of the overall variability and was mostly ascribed to BMW of the DP population, with both positive and negative values for BL, BMW, and BLW of the DS population of *G. hirsutum*. PC5 accounted for 5.1% of the overall variability, and mainly revealed the differences in LP and BUW of the DP population with their positive and negative contributions for LP of the DS population. PC6 accounted for 4.7% of overall variability, and mainly revealed the differences in LP and BW of the ML population with their positive and negative values for BW and LP of both the DS and MC populations. PC7 accounted for 4.3% of overall variability, and mainly revealed the differences in LP and BUW of the MC and DS populations, with positive and negative values for BW of the DS population. In the current study, the results of the first three PCs demonstrated a high degree of variation, indicating that inter-varietal breeding could boost cotton production.

Table 5. The correlation matrix's eigenvectors and eigenanalysis for the seven performance-related principal components of four F₂ populations.

Population	Variable	Principal Components						
		PC1	PC2	PC3	PC4	PC5	PC6	PC7
ML	BUW	0.39	0.13	−0.07	0.08	−0.04	−0.27	−0.11
	BMW	0.45	0.13	−0.08	0.07	−0.08	−0.16	−0.11
	BLW	0.45	0.13	−0.07	0.09	−0.07	−0.16	−0.08
	BL	0.36	0.04	−0.08	0.01	−0.21	0.16	0.03
	BW	0.31	0.07	−0.04	0.09	0.21	0.31	0.13
	LP	0.25	0.02	−0.03	0.05	0.13	0.57	0.09
MC	BUW	0.07	−0.16	−0.35	−0.18	0.22	−0.16	0.05
	BMW	0.03	−0.20	−0.45	−0.18	0.06	−0.01	0.13
	BLW	0.01	−0.17	−0.46	−0.18	0.08	0.01	0.08
	BL	−0.04	−0.15	−0.28	−0.19	−0.22	0.22	−0.10
	BW	−0.02	−0.11	−0.22	−0.11	−0.17	−0.07	−0.19
	LP	0.03	0.01	0.02	−0.02	−0.13	−0.33	0.79
DP	BUW	0.10	−0.32	0.07	0.16	0.30	0.08	0.07
	BMW	0.04	−0.39	0.05	0.31	0.12	−0.09	−0.042
	BLW	0.04	−0.40	0.04	0.29	−0.04	−0.09	0.03
	BL	0.05	−0.39	0.08	0.206	−0.13	0.02	0.05
	BW	0.07	−0.30	0.03	0.18	−0.25	0.05	−0.13
	LP	0.05	0.06	0.07	0.005	0.65	−0.09	−0.05
DS	BUW	0.15	−0.04	0.22	−0.177	−0.03	−0.02	0.30
	BMW	0.16	−0.18	0.28	−0.369	−0.06	0.05	0.001
	BLW	0.17	−0.19	0.28	−0.367	−0.04	0.06	0.01
	BL	0.08	−0.14	0.19	−0.38	0.03	0.13	−0.06
	BW	0.10	−0.11	0.12	−0.224	−0.01	−0.34	−0.31
	LP	−0.04	0.10	−0.01	0.157	−0.30	0.21	0.07
Eigenvalue		3.62	2.92	2.68	2.306	1.23	1.13	1.04
Variability (%)		15.10	12.20	11.20	9.60	5.10	4.70	4.30
Cumulative percentage (%)		15.10	27.30	38.50	48.10	53.20	57.90	62.20

Principal component 1 (PC1); Principal component 2 (PC2); Principal component 3 (PC3); Principal component 4 (PC4); Principal component 5 (PC5); Principal component 6 (PC6); Principal component 7 (PC7).

3.5. Biplot Analysis

These traits were superimposed as vectors onto the biplot, with the relative lengths of the vectors representing the relative degree of variance in each factor. The extant variant reserved from the origin showed higher variance and were found to be less correlated with other variants. The traits BW, BL, BUW, BMW, and BLW of the MC population; BW, BL, BUW, BMW, and BLW of the DP population; BW, BL, BUW, BMW, and BLW of the DS population; BW, BL, BUW, BMW, BLW, and LP of the MS population, respectively, were largely contributed with a high degree of variability, whereas, LP of the MC population, and LP of the DP population, represented the lowest variability. The quality trait, LP of the DS population, was not in a desirable direction (Figure 3). Meanwhile, a PCA plot was drawn for the four populations cumulatively, in which the agronomic traits BUW, BLW, BMW, and BL were contributed with a higher degree of variability. The economic traits, such as BW and LP, were negatively correlated to the agronomic traits in group performance.

Figure 3. Principal components biplot for four F_2 populations.

4. Discussion

The traits of BW and LP are complex in nature owing to their association with many other physiological attributes. Previous breeding efforts for high fiber quality traits in *G. hirsutum* remain successful. However, the *G. arboreum* germplasm has received little attention in cotton breeding programs, although it has highly significant genotypic effects on agronomic boll morphological traits and economic yield-related traits. To analyze the genetic basis for variation in BW, LP, and boll size-related traits, and their genetic architecture, three lines of *G. hirsutum* (Dixie king, Suyuan 04-44, Pamuk) and three lines of *G. arboreum* (Mei Zhongmian, Chimu Heizi, L-02292-3) were selected based on their BW and other traits. Furthermore, four F_2 populations were constructed, with Mei Zhongmian,

Chimu Heizi, and Dixie king used as female parents owing to their larger BW, while the male parents, L-02292-3, Suyuan 04-44, and Pamuk were the smaller BW lines of F_2 populations under investigation. Therefore, significant phenotypic differences were recorded in cross combinations of distantly related parents. Many selected traits showed normally distributed patterns in the four F_2 populations, which indicates the quantitative nature of the traits under consideration.

A wide range of mean values were found for the traits under observation. Mean BW ranged from 2.38 to 4.05 g in *G. arboreum* and 4.49 to 5.12 g in *G. hirsutum* (Table 3). Correlation analysis helps in the determination of associations between single traits, while multivariate analysis provides accurate measurements of similarities and differences, and determines the relative proportions of variability contributing toward total variance of the germplasm [24–29]. However, these approaches cannot evaluate the genetic factors responsible for a wide range of genetic consequences among targeted traits and their components [30,31]. Wu et al. [29] and McCarty et al. [30] investigated the advancement of conditional analysis on previous techniques to assess the contributions of within-boll yield components and genetic variability on yield-related traits. BW was found to be positively correlated with LP. Krishnarao and Mary [32] also addressed the main contributions of BW, and found strong negative correlations with LP, indicating that genotypes with higher BW tend to have lower LP. Our results were in line with previous studies, indicating that focusing exclusively on LP to improve efficiency could make BW more vulnerable [33]. Previous studies also revealed that improvements in LP are associated with reductions in boll size [17,33–36].

Conversely, BW and BL contributed a significant and positive association with phenotype variation, indicating that large bolls produce coarser fibers than smaller ones, owing to large bolls having greater sink strengths [37], thus resulting in fiber cell wall biosynthesis that is relatively successful and longer lasting. In addition, BL alone accounted for the largest phenotypic variance proportion and components variance [38]. These results provide the basis for the following assumption, that large bolls yield better and coarser fibers. Moreover, the grouping of individuals based on multiple agronomic traits is relatively useful for identifying genotype adaptation and improvement in traits of economic importance [39]. Furthermore, it was stated that biplot diagrams are foremost components in genotype selection, combining desirable traits for use in breeding programs [40]. The genotypes present in the four quadrants of the graph indicate a high source of variation and a desirable combination of traits. Principal component cluster analysis showed significant inter- and intragroup variation. Multivariate analysis is a good system for evaluating the cotton germplasm, with a critical role in identifying genotypes for biotic and abiotic stress resistance and crop yield improvement.

5. Conclusions

The contribution of *G. hirsutum* was significantly greater than that of *G. arboreum* for yield and yield-associated traits. Four F_2 populations were derived in the current study by intraspecific hybridization among two *G. hirsutum* and two *G. arboreum* lines selected on the basis of their boll size. Distinct phenotypic and genotypic correlations were reported between boll morphological, BW, and lint yield components. Furthermore, useful recombinants can be used for future breeding programs due to the lack of interaction between different traits. These findings suggest that upland cotton and breeding lines from breeding populations may possess the simultaneous genetic potential of multiple boll yield components. In comparison with distinct single-trait analysis, multivariate analysis is a good system for evaluating cotton germplasm, with a critical role in identifying genotypes of better yield and quality. Agronomical traits can provide a general representation of a population's relationship with its environment. Moreover, natural varieties may be used to transfer particular genes for various characteristics to improve cotton yield. The investigated populations might be useful in future crop improvement programs.

Author Contributions: X.D. and Z.P. developed the researched the concept. Z.I., H.G. and M.S.I. collected data and conducted formal analysis. Z.I., W.N., T.N. and Z.M. helped with statistical analysis of data. Z.I., Z.P., X.D. and D.H. helped to write the original draft of the manuscript and make revisions, T.N. and M.S.I. contributed to the review and editing processes; S.F. assisted in the literature search and writing; X.D., D.H. and H.G. reviewed the manuscript; S.F., A.M.E.-S. and A.G. provided technical expertise to streamline the idea; X.D. reviewed and supervised manuscript writing and also obtained the funding for the publication. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by funding from the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-ICR-2021-01).

Acknowledgments: We are thankful to the National mid-term gene bank for cotton for providing experimental material.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Huang, G.; Wu, Z.; Percy, R.G.; Bai, M.; Li, Y.; Frelichowski, J.E.; Hu, J.; Wang, K.; Yu, J.Z.; Zhu, Y. Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat. Genet.* **2020**, *52*, 516–524. [[CrossRef](#)] [[PubMed](#)]
- Wang, K.; Wendel, J.F.; Hua, J. Designations for individual genomes and chromosomes in *Gossypium*. *J. Cotton Res.* **2018**, *1*, 3. [[CrossRef](#)]
- Gallagher, J.P.; Grover, C.; Rex, K.; Moran, M.; Wendel, J.F. A New Species of Cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). *Syst. Bot.* **2017**, *42*, 115–123. [[CrossRef](#)]
- Hu, Y.; Chen, J.; Fang, L.; Zhang, Z.; Ma, W.; Niu, Y.; Ju, L.; Deng, J.; Zhao, T.; Lian, J.; et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* **2019**, *51*, 739–748. [[CrossRef](#)]
- Chao, F.; Ma, Y.; Wu, S.; Liu, Z.; Wang, Z.; Yang, R.; Hu, G.; Zhou, Z.; Yu, H.; Zhang, M.; et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **2017**, *18*, 161. [[CrossRef](#)]
- Chen, Z.J.; Scheffler, B.; Dennis, E.; Triplett, B.A.; Zhang, T.; Guo, W.; Chen, X.; Stelly, D.; Rabinowicz, P.D.; Town, C.D.; et al. Toward Sequencing Cotton (*Gossypium*) Genomes: Figure. *Plant Physiol.* **2007**, *145*, 1303–1310. [[CrossRef](#)]
- Wendel, J.; Brubaker, C.; Seelanan, T. The Origin and Evolution of *Gossypium*. In *Physiology of Cotton*; Stewart, J., Oosterhuis, D., Heitholt, J., Mauney, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2010.
- Ma, C.; Mamat, S.; Yao, J.; Isak, G. Spatio-temporal changes of cotton production in China from 1950 to 2015. *Acta Geogr. Sin.* **2020**, *75*, 1699–1710.
- Mehetre, S.; Mehetre, S.S.; Aher, A.R.; Gawande, V.L.; Patil, V.R.; Mokate, A.S. Induced polyploidy in *Gossypium*: A tool to overcome interspecific incompatibility of cultivated tetraploid and diploid cottons. *Curr. Sci.* **2003**, *84*, 1510–1512.
- Liu, D.; Guo, X.; Lin, Z.; Nie, Y.; Zhang, X. Genetic Diversity of Asian Cotton (*Gossypium arboreum* L.) in China Evaluated by Microsatellite Analysis. *Genet. Resour. Crop. Evol.* **2006**, *53*, 1145–1152. [[CrossRef](#)]
- Campbell, B.T.; Williams, V.E.; Park, W. Using molecular markers and field performance data to characterize the Pee Dee cotton germplasm resources. *Euphytica* **2009**, *169*, 285–301. [[CrossRef](#)]
- Tyagi, P.; Gore, M.A.; Bowman, D.T.; Campbell, B.T.; Udall, J.A.; Kuraparthi, V. Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). *Theor. Appl. Genet.* **2014**, *127*, 283–295. [[CrossRef](#)] [[PubMed](#)]
- Yang, Z.; Ge, X.; Yang, Z.; Qin, W.; Sun, G.; Wang, Z.; Li, Z.; Liu, J.; Wu, J.; Wang, Y.; et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* **2019**, *10*, 2989. [[CrossRef](#)] [[PubMed](#)]
- Ma, Z.; He, S.; Wang, X.; Sun, J.; Zhang, Y.; Zhang, G.; Wu, L.; Li, Z.; Liu, Z.; Sun, G.; et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* **2018**, *50*, 803–813. [[CrossRef](#)] [[PubMed](#)]
- Nie, J.-J.; Yuan, Y.-C.; Qin, D.-L.; Liu, Y.-H.; Wang, S.-L.; Li, J.-P.; Zhang, M.-L.; Zhao, N.; Guo, W.-J.; Qi, J.; et al. Spatial distribution of bolls affects yield formation in different genotypes of Bt cotton varieties. *J. Integr. Agric.* **2019**, *18*, 2492–2504. [[CrossRef](#)]
- Culp, T.W.; Harrell, D.C. Influence of Lint Percentage, Boll Size, and Seed Size on Lint Yield of Upland Cotton with High Fiber Strength. *Crop Sci.* **1975**, *15*, 741–746. [[CrossRef](#)]
- Zeng, L.; Meredith, W.R., Jr. Associations among lint yield, yield components, and fiber properties in an introgressed population of cotton. *Crop Sci.* **2009**, *49*, 1647–1654. [[CrossRef](#)]
- Worley, S.; Culp, T.W.; Harrell, D.C. The relative contributions of yield components to lint yield of upland cotton, *Gossypium hirsutum* L. *Euphytica* **1974**, *23*, 399–403. [[CrossRef](#)]
- Sun, S.; Li, C.; Chee, P.W.; Paterson, A.H.; Jiang, Y.; Xu, R.; Robertson, J.S.; Adhikari, J.; Shehzad, T. Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 195–207. [[CrossRef](#)]

20. Zhang, S.; Wang, T.; Liu, Q.; Gao, X.; Zhu, X.; Zhang, T.; Zhou, B. Quantitative trait locus analysis of boll-related traits in an intraspecific population of *Gossypium hirsutum*. *Euphytica* **2014**, *203*, 121–144. [\[CrossRef\]](#)
21. Majeedano, M.S.; Ahsaan, M.Z.; Somroo, A.W.; Channa, A.R. Heritability and correlation estimates for some yield traits of *Gossypium hirsutum*. *Am. Res. Thoughts* **2014**, *1*, 781–790.
22. Fruchterman, T.M.J.; Reingold, E.M. Graph drawing by force-directed placement. *Softw. Pract. Exp.* **1991**, *21*, 1129–1164. [\[CrossRef\]](#)
23. Benildo, G. Genomic and epigenomic bases of transgressive segregation—New breeding paradigm for novel plant phenotypes. *Plant Sci.* **2019**, *288*, 110213.
24. Iqbal, M.A.; Ammad, A.; Zafar, Y. Characterization of indigenous *Gossypium Arboreum*, L. Genotypes for various fiber quality traits. *Pak. J. Bot.* **2015**, *47*, 2347–2354.
25. Rezai, A.; Frey, K.J. Multivariate analysis of variation among wild oat accessions—Seed traits. *Euphytica* **1990**, *49*, 111–119. [\[CrossRef\]](#)
26. Pezzotti, M.; Falcinelli, M.; Tomassini, C.; Veronesi, F. Evaluation of an Italian germplasm collection of *Dactylis glomerata* L. using a multivariate approach. *J. Genet. Breed.* **1994**, *48*, 17–23.
27. Panthee, D.R.; Kc, R.B.; Regmi, H.N.; Subedi, P.P.; Bhattarai, S.; Dhakal, J. Diversity Analysis of Garlic (*Allium sativum* L.) Germplasms Available in Nepal Based on Morphological Characters. *Genet. Resour. Crop. Evol.* **2006**, *53*, 205–212. [\[CrossRef\]](#)
28. Flores, F.; Gutierrez, J.; Lopez, J.; Moreno, M.; Cubero, J. Multivariate analysis approach to evaluate a germplasm collection of *Hedysarum coronarium* L. *Genet. Resour. Crop. Evol.* **1997**, *44*, 545–555. [\[CrossRef\]](#)
29. Wu, J.; Jenkins, J.N.; McCarty, J.J., Jr.; Zhu, J. Genetic association of yield with its component traits in a recombinant inbred line population of cotton. *Euphytica* **2004**, *140*, 171–179. [\[CrossRef\]](#)
30. McCarty, J.C.; Wu, J.; Jenkins, J.N. Genetic association of cotton yield with its component traits in derived primitive accessions crossed by elite upland cultivars using the conditional ADAA genetic model. *Euphytica* **2008**, *161*, 337–352. [\[CrossRef\]](#)
31. McCarty, J.C.; Wu, J.; Jenkins, J.N. Primitive accession derived germplasm by cultivar crosses as sources for cotton improvement: I. Phenotypic values and variance components. *Crop Sci.* **2004**, *44*, 1226–1230. [\[CrossRef\]](#)
32. Krishnarao, K.; Mary, T. Variability, correlation and path analysis of yield and fibre traits in upland cotton. *Madras Agric. J.* **1990**, *77*, 146–151.
33. Miller, P.A.; Rawlings, J.O. Selection for Increased Lint Yield and Correlated Responses in Upland Cotton, *Gossypium hirsutum* L. *Crop Sci.* **1967**, *7*, 637–640. [\[CrossRef\]](#)
34. Bridge, R.R.; Meredith, W.R.; Chism, J.F. Comparative Performance of Obsolete Varieties and Current Varieties of Upland Cotton. *Crop Sci.* **1971**, *11*, 29–32. [\[CrossRef\]](#)
35. Ahuja, S.L.; Dhayal, L.S.; Monga, D. Performance of upland coloured cotton germplasm lines in line $\times$ tester crosses. *Euphytica* **2009**, *169*, 303–312. [\[CrossRef\]](#)
36. Zeng, L.; Meredith, W.R.; Boykin, D.L.; Taliencio, E. Evaluation of an exotic germplasm population derived from multiple crosses among *Gossypium* tetraploid species. *J. Cotton Sci.* **2007**, *11*, 118–127.
37. Xie, Z.-X.; Li, C.D.; Sun, H.C.; Zhu, J.J. Relationship between yield formation and source-sink of cotton cultivars with different boll weight. *Cotton Sci.* **2007**, *3*, 23–29.
38. Tang, F.; Xiao, W. Genetic association of within-boll yield components and boll morphological traits with fibre properties in upland cotton (*Gossypium hirsutum* L.). *Plant Breed.* **2014**, *133*, 521–529.
39. Souza, E.; Sorrells, M.E. Relationships among 70 North American Oat Germplasms: II. Cluster Analysis Using Qualitative Characters. *Crop Sci.* **1991**, *31*, 605–612. [\[CrossRef\]](#)
40. Firincioglu, H.; Erbehtas, E.; Dogruyol, L.; Mutlu, Z.; Unal, S.; Karakurt, E.; Firincioglu, H. Phenotypic variation of Autumn and Spring-sown vetch (*Vicia sativa* ssp.) populations in central Turkey. *Span. J. Agric. Res.* **2009**, *7*, 596. [\[CrossRef\]](#)